

Abstracts of Current Literature

Linguo-Statistical Studies of Siberian Languages in the USSR

Yuri A. Tambovtsev

Novosibirsk State University

PO Box 174

630058 Novosibirsk-58 USSR

Linguistische Berichte 87/83

The statistical studies of Siberian languages began in autumn 1973, so that in autumn 1983 they will celebrate their tenth anniversary. Our group of linguo-statistical studies started with phonostatistics: for practical purposes (especially for publishing) it was necessary to know the frequency of occurrence of different phonemes of Siberian native languages. The investigations were held at the Computing Centre of the Novosibirsk State University with the help of specialists in programming and speech recognition of the Laboratory of Technical Cybernetics.

The first language computed was Mansi (Vogul). The Northern and Konda dialect texts were transcribed by Mansi native speakers. Then this material, containing about half a million phonemes, was fed to a computer; the same procedure was applied to the other languages computed by the group of experimental linguistics of the Novosibirsk State University.

The group I direct has computed the following Finno-Ugric languages: Khanty (Ostyak), Udmurt (Votyak), Komi-Zyryan, Mari (Cheremis), Karelian, Finnish, Mordva (Erzya), and Saame (Lopari). In addition to the Finno-Ugric family, languages of the Turkish, Paleo-Asiatic, and Tungus-Manchurian families were computed: Khakas, Altay, Yakut, Kazakh, Ket, Eskimo, Koryak, Itelman, Nanay, Oroch, Orok, and Japanese. Every time the largest possible sample of a language was fed to the computer, but unfortunately some of them proved not to be large enough (the smallest sample contained more than 10,000 phonemes), so that the computing results of these sample should be considered preliminary. The material of these languages will be added later.

We collected the following frequency data:

- ▶ frequency of occurrence of phonemes,
- ▶ frequency of occurrence of certain phonemes in certain positions (especially in word-initial and word-final positions),
- ▶ frequency of combination of two phonemes,
- ▶ frequency of occurrence of certain dyads (combinations of two phonemes) in certain positions (especially in word-initial and word-final position),
- ▶ frequency of occurrence of triads (combinations of three phonemes).

It should be stressed that all investigations were based on phonemes but not on graphemes, since in a language one letter may or may not correspond to one sound.

Near the end of the first stage of our study – i.e., the investigations in the field of phonemic statistics – we plan to proceed with investigations in the field of lexical statistics, and then with grammatical statistical investigations of the same or enlarged material of the languages in question.

At the present time our group is collecting material on Mongol, Buryat, Tibetan, Nivkh, and some other languages of Asia. The aim of the group is to continue to compute other languages of Asia and the Far East. After that we plan to analyze and compare statistically as many languages of Siberia, Asia, and the Far East as possible.

A Computational Model for the Analysis of Arguments

Robin Cohen

Department of Computer Science

This thesis proposes a model for an argument understanding system – a natural language understanding system which processes arguments. The form of input considered is one-way communication in a conversational setting, where the speaker tries to convince the hearer of a

University of Toronto
 Toronto, CANADA M5S 1A4
*Ph.D. Thesis; Computer Systems Research
 Group Technical Report No. 151*

**Design of Natural Language Inter-
 faces: A Case Study**

Carole D. Hafner, Kurt S. Godden
 Computer Science Department
 General Motors Research Laboratories
 Warren, MI 48090-9055
Research Publication GMR-4567

particular point of view. The main contributions are: (i) a theory of expected coherent structure which limits analysis to the reconstruction of particular transmission forms; (ii) a theory of linguistic clues which assigns a functional interpretation to special words and phrases used by the speaker to indicate structure; (iii) a theory of evidence relationships which includes the demand for pragmatic analysis to accommodate beliefs not currently held. A system designed to incorporate these theories could be used to analyze the structure of arguments – the necessary first step for a hearer, before judging credibility and responding.

This paper provides an overview of the capabilities of natural language interfaces, and the design issues that must be addressed in developing a natural language system. An experimental system called DATALOG (for “database dialogue”) is described, which accepts a wide variety of English commands and questions and retrieves the answer from the user’s database. DATALOG uses a Cascaded ATN Grammar to provide efficient interaction among the different levels of knowledge representation required in a natural language system. In addition, DATALOG has the ability to give “co-operative” responses that tell the user more than the literal answer to a question. Co-operative responses are based on principles of human dialogue, which require a co-operative system to volunteer information that corrects erroneous assumptions, or that is obviously relevant to the goals of the user. Natural language technology offers a potentially valuable tool for making computer data more accessible to users; however, serious limitations of current systems in the areas of semantic coverage, dialogue-level processing, and ease of portability need to be overcome before this potential can be fulfilled.

The following technical reports are available from

Documentation Center
 USC/Information Sciences Institute
 4676 Admiralty Way
 Marina del Rey, CA 90291

The Anatomy of a Systemic Choice
William C. Mann

Report No. ISI/RR-82-104, October 1982

Systemic grammar is one of the major varieties of syntactic theory in modern linguistics. It was originally defined by Michael A.K. Halliday around 1960 and has since been developed extensively by him and others. Unlike transformational grammar, systemic grammar is oriented to the ways that language functions for its users. Systemic grammars have been used in several well-known language-processing programs and have been found to be very advantageous for computer generation of text. This report presents a framework for expressing how choices are made in systemic grammars. Formalizing the description of choice processes enriches descriptions of the syntax and semantics of languages, and it contributes to constructive models of language use. There are applications in education and computation. The framework represents the grammar as a combination of systemic syntactic description and explicit choice processes, called “choice experts.” Choice experts communicate across the boundary of the grammar to its environment, exploring an external intention to communicate. The environment’s answers lead to choices and thereby to creation of sentences and other units, tending to satisfy the intention to communication. The experts’ communicative framework includes an extension to the systemic notion of a function, in the direction of a more explicit

A Linguistic Overview of the Nigel Text Generation Grammar

William C. Mann

Report No. ISI/RS-83-9, October 1983

semantics. Choice expert processes are presented in two notations, one informal and the other formal. The information notation yields a grammar-guided conversation in English between the grammar and its environment, while the formal notation yields complete accounts of what the grammar produces given a particular circumstance and intent.

Recent text generation research resembles recent research in synthesis of vaccines. The research is designed to construct entities which previously arose naturally. This constructive approach creates practical and theoretical benefits.

Our text generation research has produced a large systemic English grammar, which is embedded in a computer program. This grammar, which is called Nigel, generates sentences. It is controlled by a semantic stratum which has been added to the basic systemic framework.

This paper describes the program, which also is called Nigel. It identifies augmentations of various precedents in the systemic framework, and it indicates the current status of the program. The paper has a dual focus. First, on Nigel's processes, it describes the methods Nigel uses to control text to fulfill a purpose by using its new semantic stratum. Second, concerning Nigel's interactions with its environment, it shows reasons why Nigel is easily embedded in a larger experimental program.

Although the paper does not focus on Nigel's syntactic scope, that its scope is non-trivial is indicated by the fact that all of the sentence and clause structures of this abstract are within that syntactic scope.

Nigel: A Systemic Grammar for Text Generation

William C. Mann,

Christian M.I.M. Matthiessen

Report No. ISI/RR-83-105, February 1983

Programming a computer to write text which meets a prior need is a challenging research task. As part of such research, Nigel, a large programmed grammar of English, has been created in the framework of systemic linguistics begun by Halliday. In addition to specifying functions and structures of English, Nigel has a novel semantic stratum which specifies the situations in which each grammatical feature should be used.

The report consists of three papers on Nigel: an introductory overview, the script of a demonstration of its use in generation, and an exposition of how Nigel relates to the systemic framework. Although the effort to develop Nigel is significant both as computer science research and as linguistic inquiry, the outlook of the report is oriented to its linguistic significance.

Relational Propositions in Discourse

William C. Mann,

Sandra A. Thompson

Report No. ISI/RR-83-115, November 1983

In addition to the propositions represented explicitly by independent clauses in a text, there are almost as many *implicit* propositions, here called relational propositions, which arise out of combinations of these clauses. The predicates of these propositions are members of a small set of general, highly recurrent relational predicates, such as "cause", "justification", and "solutionhood". Often unsignalled, these relational propositions can be shown to be the basis for other inferences and to function as elements of communicative acts. Examining two natural texts, we see that the relational propositions involve every clause, and that they occur in a pattern of propositions which connects all of the clauses together. This examination also shows how the relational propositions are essential to the functioning of the text.

Extending Grammars to New Domains

Jane J. Robinson

Report No. ISI/RR-83-123, January 1984

This is the report of an undertaking to extend and adapt an existing grammar, DIAGRAM, to provide the syntactic analysis of sentences in a new domain. DIAGRAM is an augmented phrase-structure grammar

whose rules provide a means for associating semantic and domain dependent interpretations with a syntactic analysis. An earlier version, used for the syntactic analysis and the interpretation of spoken English, covered the vocabulary and basic phrase types needed to query a static data base of information about naval ships.

The new domain in which the extended and adapted version has been tested is represented by a set of eighteen dialogues, called Helper dialogues, in which computer users present their problems to the operator and ask for help. Extending the syntax to cover the new words and phrase types exhibited in these sample texts raises a number of questions of general theoretical interest along with problems that can properly be construed as artifacts of the particular grammar that is being extended or of the limitations of the parsing program and the computer system in which the grammar is applied to input sentences. This report therefore can be read with both a broad and a narrow scope. The narrow scope reading is concerned with the additions and revisions that were made to DIAGRAM in order to parse the sentences in the dialogues selected from the new domain. The broad scope reading is concerned with the kinds of problems encountered in extending syntactic coverage generally and with strategies for coping with them.

The following reports are available from

Universitaet Konstanz
Informationswissenschaft
Projek TOPIC
Postfach 5560
D-7750 Konstanz 1
West Germany

A Formal Approach to the Semantics of a Frame Data Model: Extended Version

Ulrich Reimer, Udo Hahn

Bericht TOPIC-3/83, July 1983, 32 pp.

Standard knowledge representation languages (e.g., KRL, FRL) are almost exclusively characterized by syntactic specifications but are seriously lacking explicit formal semantic specification. The formal description of the frame data model outlined in this paper is given with emphasis on explicit semantic specifications applying a combination of a denotational and an axiomatic approach.

After introducing basic concepts of the frame data model, the set of semantic integrity constraints is outlined, finally leading to the specification of a set of basic operations in the frame data model. Based on an abstract data type view on knowledge representation languages, these operations completely specify the frame data model in terms of its behavioral properties. As these operations are the only means to access data in a frame data base, semantic integrity is always guaranteed.

Heuristic Text Parsing in TOPIC: Methodological Issues in a Knowledge-Based Text Condensation System

Udo Hahn, Ulrich Reimer

Bericht TOPIC-5/83, October 1983, 2d ed., 22 pp.

TOPIC, a knowledge-based system for text condensation and information management, is introduced with emphasis on its text parsing devices, which take into account specific requirements applying to the analysis of full texts, the generation of condensates (abstracting), and various interactive graphical facilities for text information management. The text parser under development consists of a word expert system operating on a frame knowledge base. Parsing heuristics referring to cohesion and coherence properties of texts are considered to support partial semantic parsing of the input texts.

Word Expert Parsing: An Approach to Text Parsing with a Distributed Lexical Grammar**Udo Hahn, Ulrich Reimer***Bericht TOPIC-6/83, November 1983, 27 pp.*

TOPIC, a knowledge-based text analysis system for automatic summarization of German language texts, will be described with regard to its parsing devices, which are based on the word expert model. The word expert parser currently under development is especially tuned to incorporate local cohesion and global coherence properties of expository texts as well as strategic requirements of variable-depth analysis of full texts. The technical description of corresponding text parsing procedures will be twofold, first regarding word experts from a declarative point of view as basic organizational units of a distributed text grammar while from a procedural perspective the word expert system is considered as a highly modularized text parser.

The following abstracts are from *Proceedings of the 21st Annual Conference of the Association for Computational Linguistics*, available for \$15 a copy from

Donald E. Walker, ACL
Artificial Intelligence Center
SRI International
Menlo Park, CA 94025 USA

Context-Freeness and the Computer Processing of Human Languages**Geoffrey K. Pullum**

Cowell College

University of California, Santa Cruz

Santa Cruz, CA 95064

Proc. ACL 1983, pp. 1-6

Context-free grammars, far from having insufficient expressive power for the description of human languages, may be overly powerful, along three dimensions: (1) weak generative capacity: there exists an interesting proper subset of the CFLs, the profligate CFLs, within which no human language appears to fall; (2) strong generative capacity: human languages can be appropriately described in terms of a proper subset of the CF-PSGs, namely those with the ECPO property; (3) time complexity: the recent controversy about the importance of a low deterministic polynomial time bound on the recognition problem for human languages is misdirected, since an appropriately restrictive theory would guarantee even more, namely a linear bound.

Factoring Recursion and Dependencies: An Aspect of Tree Adjoining Grammars (TAG) and a Comparison of Some Formal Properties of TAGs, GPSGs, PLGs, and LPGs**Aravind K. Joshi**Department of Computer and
Information Science

R. 268 Moore School

University of Pennsylvania

Philadelphia, PA 19104

Proc. ACL 1983, pp. 7-15

During the last few years there is vigorous activity in constructing highly constrained grammatical systems by eliminating the transformational component either totally or partially. There is increasing recognition of the fact that the entire range of dependencies that transformational grammars in their various incarnations have tried to account for can be satisfactorily captured by classes of rules that are *non-transformational* and at the same time highly constrained in terms of the classes of grammars and languages that they define.

In this paper, we will first briefly describe TAGs, which have the following important properties: (1) we can represent the usual transformational relations more or less directly in TAGs, (2) the power of TAGs is only slightly more than that of context-free grammars (CFGs) in what appears to be just the right way, and (3) TAGs are powerful enough to characterize dependencies (e.g., subcategorization, as in verb subcategorization, and filler-gap dependencies, as in the case of moved constituents in wh-questions) which might be at unbounded distance and nested or crossed. We will then compare some of the formal properties of TAGs, GPSGs, PLGs, and LFGs, in particular, concerning (1) the types of languages, reflecting different patterns of dependencies that can or cannot be generated by the different types of grammars, (2) the degree of free word ordering permitted by different grammars, and (3) parsing complexity of the different grammars.

Crossed Serial Dependencies: A Low-Power Parseable Extension to GPSG
Henry Thompson

Department of Artificial Intelligence
 and Program in Cognitive Science
 University of Edinburgh
 Hope Park Square, Meadow Lane
 Edinburgh EH8 9NW SCOTLAND
Proc. ACL 1983, pp. 16-21

An extension to the GPSG grammatical formalism is proposed, allowing non-terminals to consist of finite sequences of category labels, and allowing schematic variables to range over such sequences. The extension is shown to be sufficient to provide a strongly adequate grammar for crossed serial dependencies, as found for example in Dutch subordinate clauses. The structures induced for such constructions are argued to be more appropriate to data involving conjunction than some previous proposals have been. The extension is shown to be parseable by a simple extension to an existing parsing method of GPSG.

Formal Constraints on Metarules
Stuart M. Shieber, Susan U. Stucky,
Hans Uszkoreit, Jane J. Robinson

SRI International
 333 Ravenswood Avenue
 Menlo Park, CA 94025
Proc. ACL 1983, pp. 22-27

Metagrammatical formalisms that combine context-free phrase structure rules and metarules (MPS grammars) allow concise statement of generalizations about the syntax of natural languages. Unconstrained MPS grammars, unfortunately, are not computationally "safe". We evaluate several proposals for constraining them, basing our assessment on computational tractability and explanatory adequacy. We show that none of them satisfies both criteria, and suggest new directions for research on alternative metagrammaticalisms.

A Prolegomenon to Situation Semantics

David J. Israel

Bolt Beranek and Newman, Inc.
 Cambridge, MA 02238
Proc. ACL 1983, pp. 28-37

An attempt is made to prepare Computational Linguistics for Situation Semantics.

A Modal Temporal Logic for Reasoning about Change
Eric Mays

Department of Computer and
 Information Science
 Moore School of Electrical
 Engineering / D2
 University of Pennsylvania
 Philadelphia, PA 19104
Proc. ACL 1983, pp. 38-43

We examine several behaviors for query systems that become possible with the ability to represent and reason about change in data bases: queries about possible futures, queries about alternative histories, and offers of monitors as responses to queries. A modal temporal logic is developed for this purpose. A completion axiom for history is given and modelling strategies are given by example.

Providing a Unified Account of Definite Noun Phrases in Discourse

Barbara J. Grosz

Artificial Intelligence Center
 SRI International
 Menlo Park, CA 94025

Aravind K. Joshi

Department of Computer and
 Information Science

Scott Weinstein

Department of Philosophy
 University of Pennsylvania
 Philadelphia, PA 19104
Proc. ACL 1983, pp. 44-50

Linguistic theories typically assign various linguistic phenomena to one of the categories, *syntactic*, *semantic*, or *pragmatic*, as if the phenomena in each category were relatively independent of those in the others. However, various phenomena in discourse do not seem to yield comfortably to any account that is strictly a syntactic or semantic or pragmatic one. This paper focuses on particular phenomena of this sort – the use of various referring expressions such as definite noun phrases and pronouns – and examines their interaction with mechanisms used to maintain discourse coherence.

Using λ -Calculus to Represent Meanings in Logic Grammars
David Scott Warren

Computer Science Department
 SUNY at Stony Brook

This paper describes how meanings are represented in a semantic grammar for a fragment of English in the logic programming language Prolog. The conventions of Definite Clause Grammars are used. Previous work on DCGs with a semantic component has used essentially first-order formulas for representing meanings. The system de-

Stony Brook, NY 11794
Proc. ACL 1983, pp. 51-56

An Improper Treatment of Quantification in Ordinary English

Jerry R. Hobbs

SRI International
 Menlo Park, CA 94025
Proc. ACL 1983, pp. 57-63

A Foundation for Semantic Interpretation

Graeme Hirst

Department of Computer Science
 Brown University
 Providence, RI 02912
Proc. ACL 1983, pp. 64-73

TELEGRAM: A Grammar Formalism for Language Planning

Douglas E. Appelt

Artificial Intelligence Center
 SRI International
 Menlo Park, CA 94025
Proc. ACL 1983, pp. 74-78

scribed here uses formulas of the typed λ -calculus. The first section discusses general issues concerning the use of first-order logic or the λ -calculus to represent meanings. The second section describes how λ -calculus meaning representations can be constructed and manipulated directly in Prolog. This 'programmed' representation motivates a suggestion, discussed in the third section, for an extension to Prolog so that the language itself would include a mechanism for handling the λ -formulas directly.

In the currently standard ways of representing quantification in logical form, the sentence

In most democratic countries most politicians can fool most of the people on almost every issue most of the time. has 120 different readings, or quantifier scopings. Moreover, they are truly distinct, in the sense that for any two readings there is a model that satisfies one and not the other. What is needed is a logical form for such sentences that is neutral with respect to the various scoping possibilities. The approach taken here uses the notion of "typical element" of a set to produce a flat logical form of conjoined atomic predications. A treatment has been worked out only for monotone increasing determiners; some ideas about other determiners are discussed. An inferencing component capable of resolving coreference, doing coercions, and refining predicates will be assumed (but not discussed). Thus, translating the quantifier scoping problem into one of those three processes will count as a solution for the purposes of this paper.

Traditionally, translation from the parse tree representing a sentence to a semantic representation (such as frames or procedural semantics) has always been the most ad hoc part of natural language understanding (NLU) systems. However, recent advances in linguistics, most notably the system of formal semantics known as Montague semantics, suggest ways of putting NLU semantics onto a cleaner and firmer foundation. We are using a Montague-inspired approach to semantics in an integrated NLU and problem-solving system that we are building. Like Montague's our semantics are compositional by design and strongly typed, with semantic rules in one-to-one correspondence with the meaning-affecting rules of a Marcus-style parser. We have replaced Montague's semantic objects, functors and truth conditions, with the elements of the frame language Frail, and added a word sense and case slot disambiguation system. The result is a foundation for semantic interpretation that we believe to be superior to previous approaches.

Planning provides the basis for a theory of language generation that considers the communicative goals of the speaker when producing utterances. One central problem in designing a system based on such a theory is specifying the requisite linguistic knowledge in a form that interfaces well with a planning system and allows for the encoding of discourse information. The TELEGRAM (TELEological GRAMmar) system described in this paper solves this problem by annotating a unification grammar with assertions about how grammatical choices are used to achieve various goals, and by enabling the planner to augment the functional description of an utterance as it is being unified. The control structures of the planner and the grammar unifier are then merged in a manner that makes it possible for general planning to be guided by unification of a particular functional description.

An Overview of the Nigel Text Generation Grammar

William C. Mann

USC/Information Sciences Institute
4676 Admiralty Way #1101
Marina del Rey, CA 90291
Proc. ACL 1983, pp. 79-84

Research on the text generation task has led to the creation of a large systemic grammar of English, Nigel, which is embedded in a computer program. The grammar and the systemic framework have been extended by addition of a semantic stratum. The grammar generates sentences and other units under several kinds of experimental control.

This paper describes augmentations of various precedents in the systemic framework. The emphasis is on developments which control the text to fulfill a purpose, and on characteristics which make Nigel relatively easy to embed in a larger experimental program.

Automatic Recognition of Intonation Patterns

Janet B. Pierrehumbert

Bell Laboratories
Murray Hill, NJ 07974
Proc. ACL 1983, pp. 85-90

This paper is a progress report on a project in linguistically based automatic speech recognition. The domain of this project is English intonation. The system I will describe analyzes fundamental frequency contours (F0 contours) of speech in terms of the theory of melody laid out in Pierrehumbert (1980). Experiments discussed in Liberman and Pierrehumbert (1983) support the assumptions made about intonational phonetics, and an F0 synthesis program based on a precursor to the present theory is described in Pierrehumbert (1981).

A Finite-State Parser for Use in Speech Recognition

Kenneth W. Church

NE43-307
Massachusetts Institute of Technology
Cambridge, MA 02139
Proc. ACL 1983, pp. 91-97

This paper is divided into two parts. The first section motivates the application of finite-state parsing techniques at the phonetic level in order to exploit certain classes of contextual constraints. In the second section, the parsing framework is extended in order to account for 'feature spreading' (e.g., agreement and co-articulation) in a natural way.

On the Mathematical Properties of Linguistic Theories

C. Raymond Perrault

Department of Computer Science
University of Toronto
Toronto, Ont, Canada M5S 1A4
Proc. ACL 1983, pp. 98-105

Meta-theoretical results on the decidability, generative capacity, and recognition complexity of several syntactic theories are surveyed. These include context-free grammars, transformational grammars, lexical functional grammars, generalized phrase structure grammars, and tree adjunct grammars.

A Framework for Processing Partially Free Word Order

Hans Uszkoreit

Artificial Intelligence Center
SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025
Proc. ACL 1983, pp. 106-112

The partially free word order in German belongs to the class of phenomena in natural language that requires a close interaction between syntax and pragmatics. Several competing principles, which are based on syntactic and on discourse information, determine the linear order of noun phrases. A solution to problems of this sort is a prerequisite for high-quality language generation. The linguistic framework of Generalized Phrase Structure Grammar offers tools for dealing with word order variation. Some slight modifications to the framework allow for an analysis of the German data that incorporates just the right degree of interaction between syntactic and pragmatic components and that can account for conflicting ordering statements.

Sentence Disambiguation by a Shift-Reduce Parsing Technique

Stuart M. Shieber

Artificial Intelligence Center
SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025
Proc. ACL 1983, pp. 113-118

Native speakers of English show definite and consistent preferences for certain readings of syntactically ambiguous sentences. A user of a natural-language-processing system would naturally expect it to reflect the same preferences. Thus, such systems must model in some way the *linguistic performance* as well as the *linguistic competence* of the native speaker. We have developed a parsing algorithm – a variant of the LALR(1) shift-reduce algorithm – that models the preference behavior of native speakers for a range of syntactic preference phenomena reported in the psycholinguistic literature, including the recent data on lexical preferences. The algorithm yields the preferred parse determin-

Syntactic Constraints and Efficient Parsability**Robert C. Berwick**

Room 820

MIT Artificial Intelligence Laboratory

545 Technology Square

Cambridge, MA 02139

Amy S. Weinberg

Department of Linguistics, MIT

Cambridge, MA 02139

Proc. ACL 1983, pp. 119-122

istically, without building multiple parse trees and choosing among them. As a side effect, it displays appropriate behavior in processing the much discussed garden-path sentences. The parsing algorithm has been implemented and has confirmed the feasibility of our approach to the modeling of these phenomena.

A central goal of linguistic theory is to explain why natural languages are the way they are. It has often been supposed that computational considerations ought to play a role in this characterization, but rigorous arguments along these lines have been difficult to come by. In this paper we show how a key "axiom" of certain theories of grammar, Subjacency, can be explained by appealing to general restrictions on on-line parsing plus natural constraints on the rule-writing vocabulary of grammars. The explanation avoids the problems with Marcus's (1980) attempt to account for the same constraint. The argument is robust with respect to machine implementation, and thus avoids the problems that often arise when making detailed claims about parsing efficiency. It has the added virtue of unifying the functional domain of parsing certain grammatically disparate phenomena, as well as making a strong claim about the way in which the grammar is actually embedded into an on-line sentence processor.

Deterministic Parsing of Syntactic Non-fluencies**Donald Hindle**

Bell Laboratories

Murray Hill, NJ 07974

Proc. ACL 1983, pp. 123-128

It is often remarked that natural language, used naturally, is unnaturally ungrammatical. Spontaneous speech contains all manner of false starts, hesitations, and self-corrections that disrupt the well-formedness of strings. It is a mystery then, that despite this apparent wide deviation from grammatical norms, people have little difficulty understanding the non-fluent speech that is the essential medium of everyday life. And it is a still greater mystery that children can succeed in acquiring the grammar of a language on the basis of evidence provided by a mixed set of apparently grammatical and ungrammatical strings.

D-Theory: Talking about Talking About Trees**Mitchell P. Marcus, Donald Hindle,****Margaret M. Fleck**

Bell Laboratories

Murray Hill, NJ 07974

Proc. ACL 1983, pp. 129-136

Linguists, including computational linguists, have always been fond of talking about trees. In this paper we outline a theory of linguistic structure that talks about talking about trees; we call this theory *Description theory* (D-theory). While important issues must be resolved before a complete picture of D-theory emerges (and also before we can build programs that utilize it), we believe that this theory will ultimately provide a framework for explaining the syntax and semantics of natural language in a manner that is *intrinsically* computational. This paper will focus primarily on one set of motivations for this theory, those engendered by attempts to handle certain syntactic phenomena within the framework of deterministic parsing.

Parsing as Deduction**Fernando C.N. Pereira,****David H.D. Warren**

Artificial Intelligence Center

SRI International

333 Ravenswood Avenue

Menlo Park, CA 94025

Proc. ACL 1983, pp. 137-144

By exploring the relationship between parsing and deduction, a new more general view of chart parsing is obtained, which encompasses parsing for grammar formalisms based on unification, and is the basis of the Earley Deduction proof procedure for definite clauses. The efficiency of this approach for an interesting class of grammars is discussed.

Design of a Knowledge-Based Report Generator**Karen Kukich**

University of Pittsburgh

Knowledge-Based Report Generation is a technique for automatically generating natural language reports from computer data bases. It is so named because it applies knowledge-based expert systems software to the problem of text generation. The first application of the technique,

Bell Laboratories
Murray Hill, NJ 07974
Proc. ACL 1983, pp. 145-150

a system for generating natural language stock reports from a daily stock quotes data base, is partially implemented. Three fundamental principles of the technique are its use of domain-specific semantic and linguistic knowledge, its use of macro-level semantic and linguistic constructs (such as whole messages, a phrasal lexicon, and a sentence-combining grammar), and its production system approach to knowledge representation.

Menu-Based Natural Language Understanding

Harry R. Tennant, Kenneth M. Ross, Richard M. Saenz, Craig W. Thompson, James R. Miller

Computer Science Laboratory
Central Research Laboratories
Texas Instruments Incorporated
Dallas, TX

Proc. ACL 1983, pp. 151-158

This paper describes the NLMenu System, a menu-based natural language understanding system. Rather than requiring the user to type his input to the system, input to NLMenu is made by selecting items from a set of dynamically changing menus. Active menus and items are determined by a predictive left corner parser that accesses a semantic grammar and lexicon. The advantage of this approach is that all inputs to the NLMenu System can be understood, thus giving a 0% failure rate. A companion system that can automatically generate interfaces to relational data bases is also discussed.

Knowledge Structures in UC, The UNIX* Consultant

David N. Chin

Division of Computer Science
Department of EECS
University of California, Berkeley,
Berkeley, CA 94720

Proc. ACL, 1983, pp. 159-163

**Trademark of Bell Laboratories*

The knowledge structures implemented in UC, the UNIX Consultant, are sufficient for UC to reply to a large range of user queries in the domain of the UNIX operating system. This paper describes how these knowledge structures are used in the natural language tasks of parsing, inference, planning, goal detection, and generation, and how they are organized to enable efficient access even with the large data base of an expert system. The structuring of knowledge to provide direct answers to common queries and the high usability and efficiency of knowledge structures allow UC to hold an interactive conversation with a user.

Discourse Pragmatics and Ellipsis Resolution in Task-Oriented Natural Language Interface

Jaime G. Carbonell

Computer Science Department
Carnegie-Mellon University
Pittsburgh, PA 15213

Proc. ACL 1983, pp. 164-168

This paper reviews discourse phenomena that occur frequently in task-oriented man-machine dialogs, reporting on an empirical study that demonstrates the necessity of handling ellipsis, anaphora, extra-grammaticality, inter-sentential metalanguage, and other abbreviatory devices in order to achieve convivial user interaction. Invariably, users prefer to generate terse or fragmentary utterances instead of longer, more complete "stand-alone" expressions, even when given clear instructions to the contrary. The XCALIBUR expert system interface is designed to meet these needs, including generalized ellipsis resolution by means of a rule-based caseframe method superior to previous semantic grammar approaches.

